Environmental Toxicology and Chemistry, Vol. 19, No. 4(2), pp. 1066–1075, 2000

Printed in the USA
0730-7268/00 $9.00 + .00

BOTTOM-UP AND TOP-DOWN APPROACHES TO ASSESS MULTIPLE STRESSORS OVER LARGE GEOGRAPHIC AREAS

SCOTT D. DYER,*† CHARLOTTE WHITE-HULL,† GREGORY J. CARR,† ERIC P. SMITH,‡ and XINHAO WANG§
†The Procter & Gamble Company, Cincinnati, Ohio 45253, USA
‡Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, USA
§University of Cincinnati, Cincinnati, Ohio 45221, USA

(*Received* 5 *January* 1999; *Accepted* 26 *July* 1999)

Abstract—The relationship of multiple factors, such as instream habitat, drainage area, gradient, cumulative effluent, conventional pollutants, and chemical mixtures, to fish communities was explored at the subbasin, basin, and state level within the state of Ohio, USA. Two approaches were used: bottom-up, which focused on subbasin- and basin-level relationships within the Great Miami River, Ohio, and top-down, focusing on relationships across the entire state. Data were provided by the Ohio Environmental Protection Agency and the U.S. Environmental Protection Agency. These data were integrated via a geographical information system. Multiple linear regression was used to determine the strength of stressor–response relationships. The greatest amount of variation of the index of biotic integrity (IBI) and selected metrics was addressed at the subbasin level, followed by the basin and state level, respectively. Overall, habitat factors were the best predictors and positively related to the IBI and number of fish species. Chemical factors, such as cumulative effluent, metals, ammonia, and biochemical oxygen demand, were consistently observed as negative, moderating factors for IBI and fish taxa richness and were the best predictors of the percent of fish observed with deformities, fin erosions, lesions, and tumors.

Keywords—Mixtures Geographic information system Index Effluent Biotic integrity

INTRODUCTION

The presence, abundance, diversity, and distribution of fish communities is dependent on their ability to respond to a variety of physical and chemical factors to which they are exposed. Although the purpose of risk assessment is to estimate the probability of an adverse effect of a stressor (e.g., chemicals) to a specified receiving compartment (e.g., fish community), risk assessment is typically limited to a single stressor. Therefore, risk of a stressor that incorporates the interaction of other, diverse stressors should more accurately reflect reality and be verifiable. Although this is a long-term goal in risk assessment, methods for analyzing and assessing the interaction of multiple stressors, such as degraded instream habitat, ecological constraints imposed by river size and morphology, and exposure to effluent and chemical mixtures, need to be developed or refined. Further, the relationships that stressors impose on receiving water biota (e.g., fish communities) may be scale dependent. This is of particular importance to lotic systems, where the potency of a stressor may not be directly proportional to its proximity to fish sampling stations, but may be distant or a function of the accumulation of similar stressors (e.g., chemical mixtures, cumulative exposure to effluent). Before risk of a particular stressor or multiple stressors can be estimated, exploration of their relationships to a receiving community at different geographic scales is needed.

The interaction of multiple stressors with fish communities in Ohio, USA, were assessed using two different approaches: bottom-up and top-down. The bottom-up approach focused on developing an understanding of the key factors driving the receiving water ecology at a basin or subbasin level, with emphasis given to the Great Miami River, located in the southwest corner of Ohio. The top-down approach attempted to draw relationships of stressors to fish communities across the entire state. In essence, the effects of analyzing data sets at three different levels of scale (subbasin, basin, and state) were investigated.

METHODS

Data sources

Water chemistry data. Ambient water chemistry data for the entire state of Ohio were extracted from STORET, a comprehensive data collection and reporting database describing surface and groundwater quality for North American waterways [1]. Each monitoring site, or station, in the database contains water chemistry data keyed by parameter (e.g., NH3, metals) and other descriptive information. Data extracted for the study included station name, agency, river reach number, latitude and longitude, and measured chemical concentrations. Parameters extracted for the study are presented in Table 1.

Water chemistry data were retrieved for years 1990 through 1996 to overlap the time frame in which the entire state was monitored for its biotic integrity. The log-transformed median (M) and 90th percentile (90) concentrations for each water chemistry parameter per station were determined.

Mixture toxicity. The derivation of the total toxic load of contaminants at each sampling site was based on the concept of effects addition. Effects benchmarks for each chemical considered were based on established U.S. Environmental Protection Agency water quality criteria. Algorithms used to derive toxic units were summarized in Dyer et al. [2]. Metals and ammonia were the primary contributors to the mixture

* To whom correspondence may be addressed (dyer.sd@pg.com).
Presented at the American Society for Testing and Materials–U.S. Environmental Protection Agency–Society of Environmental Toxicology and Chemistry Symposium on Ecosystem Vulnerability, Seattle, Washington, USA, August 17–20, 1998.

Table 1. List of water chemistry, stream habitat, and fish metrics that were used in the study. Water chemistry information was extracted from the U.S. Environmental Protection Agency's database STORET, and habitat and fish parameters were obtained from the Ohio Environmental Protection Agency. M and 90 refer to the median and 90th percentile concentrations per site

Parameters	Units or scale	STORET no.
Water chemistry		
Alkalinity, total (Alktot M or 90)	mg/L as $CaCO_3$	410
Aluminum, total (Altot M or 90)	μg/L as A1	1105
Cadmium, total (Cdtot M or 90)	μg/L as Cd	1027
Carbon, total organic (TOC M or 90)	mg/L as C	680
Carbonaceous biochemical oxygen demand (CBOD), 5 d, 20°C (BOD M or 90)	mg/L	80082
Cumulative percent municipal wastewater treatment plant effluent at low + mean river flows (%Effl L or M)	Percent	
Copper, total (Cutot M or 90)	μg/L as Cu	1042
Dissolved oxygen (DO M or 90)	mg O_2/L	300
Hardness, total (Hard M or 90)	mg/L as $CaCO_3$	900
Lead, total (Pbtot M or 90)	μg/L as Pb	1051
Manganese, total (Mntot M or 90)	μg/L as Mn	1055
Nickel, total (Nitot M or 90)	μg/L as Ni	1067
Nitrogen, ammonia (NH3 or 90)	Total (mg/L) as N	610
pH (pH M or 90)	Standard units	400
Phosphorus, total (TotP M or 90)	mg/L as P	665
Residue, total nonfilterable (TSS M or 90)	mg/L	530
Selenium, total (Setot M or 90)	μg/L as Se	1147
Silver, total (Agtot M or 90)	μg/L as Ag	1077
Total toxic units (TotTUs M or 90)	TUs	
Zinc, total (Zntot M or 90)	μg/L as Zn	1092
Habitat		
Substrate	0–20	
Instream cover (Cover)	0–20	
Channel quality (Channel)	0–20	
Riparian/erosion (Riparian)	0–10	
Pool or Riffle	0–20	
Qualitative habitat evaluation index (QHEI)	0–100	
Gradient	ft/mi	
Drainage Area	mi^2	
Biological		
No. fish species (Species)		
Percent deformities, fin erosions, lesions, or tumor anomalies (DELT)	0–100	
Index of biotic integrity (IBI)	12–60	

evaluation, because too few organic contaminant data were available.

Habitat and fish data. Habitat and fish data for 1990 to 1996 were provided by the Ohio Environmental Protection Agency, Columbus, Ohio, USA. During this period each of the state's major watersheds was sampled. Typically, the Ohio Environmental Protection Agency samples each watershed once every five years. Table 1 includes all of the metrics and indices used in this study.

A description of the habitat metrics and their use in deriving the qualitative habitat evaluation index (QHEI) for Ohio is provided by Rankin [3]. Briefly, the QHEI is derived from six metrics (capitalized): Substrate, instream cover (Cover), channel quality (Channel), riparian/erosion (Riparian), Pool/Riffle, and Gradient. In addition to the six metrics, drainage area was included as a habitat variable. Scores for the first six metrics and raw data for gradient (ft/mi) and drainage area (mi^2) were used. The sum of the six metrics is 100, with 100 indicating the highest quality habitat. Waters with QHEI scores of less than 45 can be considered limiting to aquatic life, whereas waters with scores greater than 60 are considered good habitats [4].

Information on the status of fish communities was obtained using metrics corresponding to the index of biotic integrity (IBI) [4]. The IBI is derived from 12 metrics, scored from 1 to 5, with 5 being the highest, most favorable value. Fish were collected via electroshocking methods. A brief description for 9 of the 12 metrics used in this study follows. The total number of species (Species) refers to indigenous species only and excludes introduced species. A high number of darter species (Darters) is an indicator of good water quality and habitat conditions, particularly for headwaters and sites suitable for wading. The percent of round-bodied suckers (Rdsuckpc) is substituted for Darters in the derivation of the IBI in sites where boat sampling is required. The number of sunfish species (Sunfish) refers to all taxa within the family Centrarchidae, excluding *Micropterus* spp. and *Lepomis microlophus.* The number of sucker species (Suckers) refers to all catostomid species inhabiting wading and boat sampling sites. The numbers of intolerant species (Intols, wading and boat sites) are used to distinguish streams with high water and habitat quality. Examples of species included in these categories are: blue sucker, river redhorse, river chub, silver shiner, stonecat, brindled madtom, and variegate darter. The percent tolerant species (Tolperc) includes central mudminnow, white sucker, carp, creek chub, bluntnose and fathead minnows, green sunfish, and yellow and brown bullheads. The percent omnivores (Omnivor, e.g., threadfin shad, carpsuckers, carp, and fathead minnow) is an indicator of increasing environmental degradation due to disruption of the food chain. The percent of top carnivores (Topcarn, e.g., American eel, rock bass, and largemouth and smallmouth bass) is used to designate integrity in the upper functional levels of the fish community. The proportion of individuals with deformities, eroded fins, lesions, tumors, and other abnormalities (DELTs) is used as an indicator of severe disturbances to the water quality and habitat of the receiving water. Typically, increased incidences of DELTs are found downstream of municipal and industrial wastewater discharges [4]. Because of space constraints within this manuscript, use of two metrics (Species and DELTs) and the IBI are discussed in the bottom-up approach, whereas all the other metrics mentioned above were used in the top-down approach.

River network data. River characteristics data were extracted from the U.S. Environmental Protection Agency's Reach File 1 (RF1) [5], a hydrologic database of the surface waters of the United Sstates in geographic information system line coverage format. In this database, river systems are represented as uniquely coded stream segments (i.e., reaches), which have been hydrologically linked to perform navigation for modeling applications and data retrieval. The database contains 68,000 records for the entire United States, encompassing perennial streams at a scale of 1:500,000. Extracted data included reach number, stream name, type, order, and length for all river reaches.

Point source data. Wastewater treatment plant (WWTP) location data were extracted from the U.S. Environmental Protection Agency's Needs Survey [6] and Permit Compliance System databases (PCS) [7]. The PCS database manages discharge monitoring data and permit compliance for the National Pollutant Discharge Elimination System. The database contains more than 19 million records describing permitted municipal and industrial dischargers. The 1988 U.S. Environmental Protection Agency Needs Survey database describes approximately 11,500 municipal WWTPs. The biannual survey is designed to provide the U.S. Congress with an inventory of existing WWTPs and estimates of future treatment plant needs. The 1988 survey was used because it was the most comprehensive and accurate survey at study initiation. Data extracted for this study included latitude and longitude, and permit number for each municipal discharger. Duplicate information (e.g., National Pollutant Discharge Elimination System number) from Needs Survey and PCS databases was eliminated. Industrial point sources were also excluded from the study. A total of 657 Ohio facilities discharging to RF1 river reaches were included in the final analysis.

Data integration

Biological, chemical, and habitat monitoring sites rarely occurred at the exact same latitude and longitude. In order to compare data from each of these data sets in a spatially relevant manner, an aggregation scheme was required. We established the spatial relationship through the following steps with ARC/INFO version 7.0.4 and ArcView GIS 3.0, commercial geographic information system software developed by the Environmental Systems Research Institute (Redlands, CA, USA).

River network. The baseline map for Ohio rivers was the RF1 file. However, the data do not have network features, which are essential for establishing river order, upstream–downstream relationship, and measuring river length from the head water to the mouth. Further, RF1 river reaches often cross WWTP discharge points, which poses a major hydrological and chemical problem because the flows of many streams and rivers in Ohio are dominated by effluents at mean and/or low flows. To resolve this, ARC/INFO network functions were used to create a river network from the RF1 line file. In a network, line segments with the same river name are grouped into routes. The total length of a river can then be calculated by adding lines in the route. Therefore, river mileage at any point along a river can be calculated and retrieved. River routes were further divided into segments based on the river hydrologic features and the location of WWTP discharge points. A river route was broken into segments based on three criteria: river confluence, WWTP discharge point on the river, or 2-mi maximum. Segments less than 30 m long using these criteria were combined with the next downstream segment. Implementation of this segmentation scheme across the state resulted in 5,879 segments, with an average segment length of approximately 1.8 mi. Each segment was assigned a unique identifier, which was used to compare all monitored data.

Point files. Biological, habitat, and water chemistry files were converted to ARC/INFO point coverages using ARC/INFO's geocoding function to assign each site a location based on its longitude and latitude values. The point coverages were then projected to the Albers conic equal-area projection, the projection used for the river network data.

Spatial aggregation. The ARC/INFO spatial overlay functions were used to assign segment numbers to each point file. To accomplish this, the nearest distance from each monitoring site to a river route was calculated and the segment number on that route was assigned to the site. Monitoring sites located on non-RF1 (i.e., smaller-order) streams were not incorporated into the river routing unless located within 1 mi of a larger river segment included in the routing. For each data type having more than one point per river segment, mean values per metric were obtained from all points.

Cumulative effluent. Cumulative percent WWTP effluent serves as a surrogate for persistent wastewater contributions to receiving water quality. To perform this calculation, mean and critical low (7Q10) stream flow for each river segment was first extracted from the U.S. Geological Services Gage database [5]. Percent cumulative effluent was calculated as the ratio of WWTP flow to stream flow for headwater segments. For all other segments, WWTP flow included not only contributions from facilities on that river segment, but also contributions from facilities upstream (e.g., main stem, tributaries) of that segment.

Bottom-up analyses

Study area: Great Miami River. The Great Miami River watershed (GMR), located in southwestern Ohio, is adjacent to the Cincinnati and Dayton metropolitan area (Fig. 1). The GMR drains an area of 5,385 mi^2 and has a main stem length of 170 mi with an average gradient of 3.9 ft/mi [8]. The majority of the GMR lies within the Eastern Corn Belt Plains ecoregion, whereas the lower portion flows through the Interior Plateau. Major tributaries to the GMR include Twin Creek, Mad, and Stillwater rivers. Upstream of Dayton significant numbers of reaches of the GMR have been designated as exceptional warm-water habitat [9]. Other exceptional warm-water habitat–classified waters in the basin include the Stillwater River and Twin Creek [10]. All other waters in the GMR basin are generally classified as warm-water habitat. An IBI score of >50 is used to derive an exceptional warm-water habitat classification, whereas a score of between 40 and 50 is required for a warm-water habitat classification.

Regression analysis. Viewing the variance of water chemistry, habitat, and biological data on maps via ARCVIEW indicated that spatial clusters of characteristics (subbasins) occurred, which may be important in statistical analysis and in-

Fig. 1. Location of the Great Miami River (dark gray) and major urban areas in Ohio, USA.

Table 2. Basin codes and names used for top-down approach for Ohio, USA

Basin code	Name
01	Hocking River
02	Scioto River
03	Grand River
04	Maumee River
05	Sandusky River
06	Central Ohio River tributaries
07	Ashtabula River
08	Little Beaver Creek
09	Southeastern Ohio River tributaries
10	Southwestern Ohio River tributaries
11	Little Miami River
12	Huron River
13	Rocky River
14	Great Miami River
15	Chagrin River
16	Portage River
17	Muskingum River
18	Mahoning River
19	Cuyahoga River
20	Black River
21	Vermilion River
22	Wabash River
23	Mill Creek

terpretation. Hence, statistical analyses were conducted on the entire GMR as well as five subbasins: Upper GMR, Lower GMR, Mad River, Stillwater River, and Twin Creek and other tributaries (Fig. 2).

Multiple linear regression models to predict IBI, Species, and DELTs were investigated through a combination of all possible regressions and forward selection [11]. All possible regressions was used first to determine the three best (maximum squared multiple correlation, R^2) one-variable models,

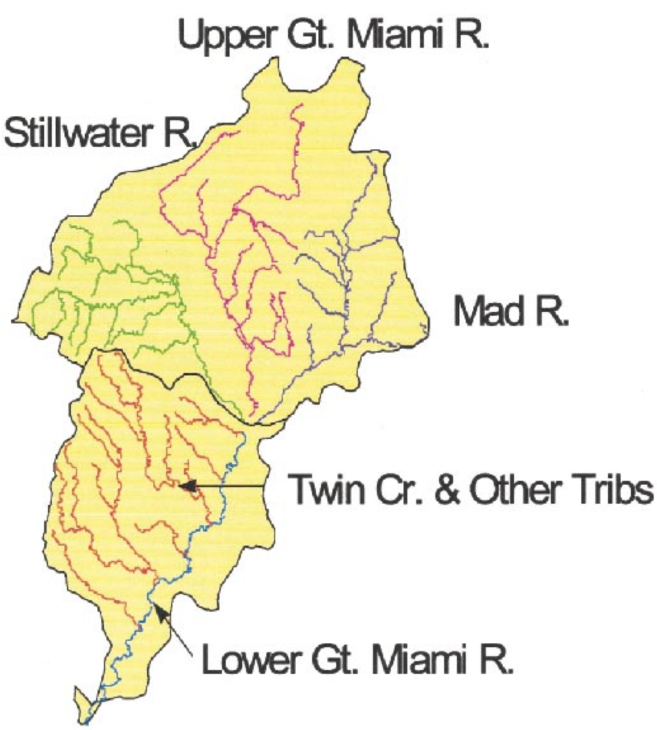
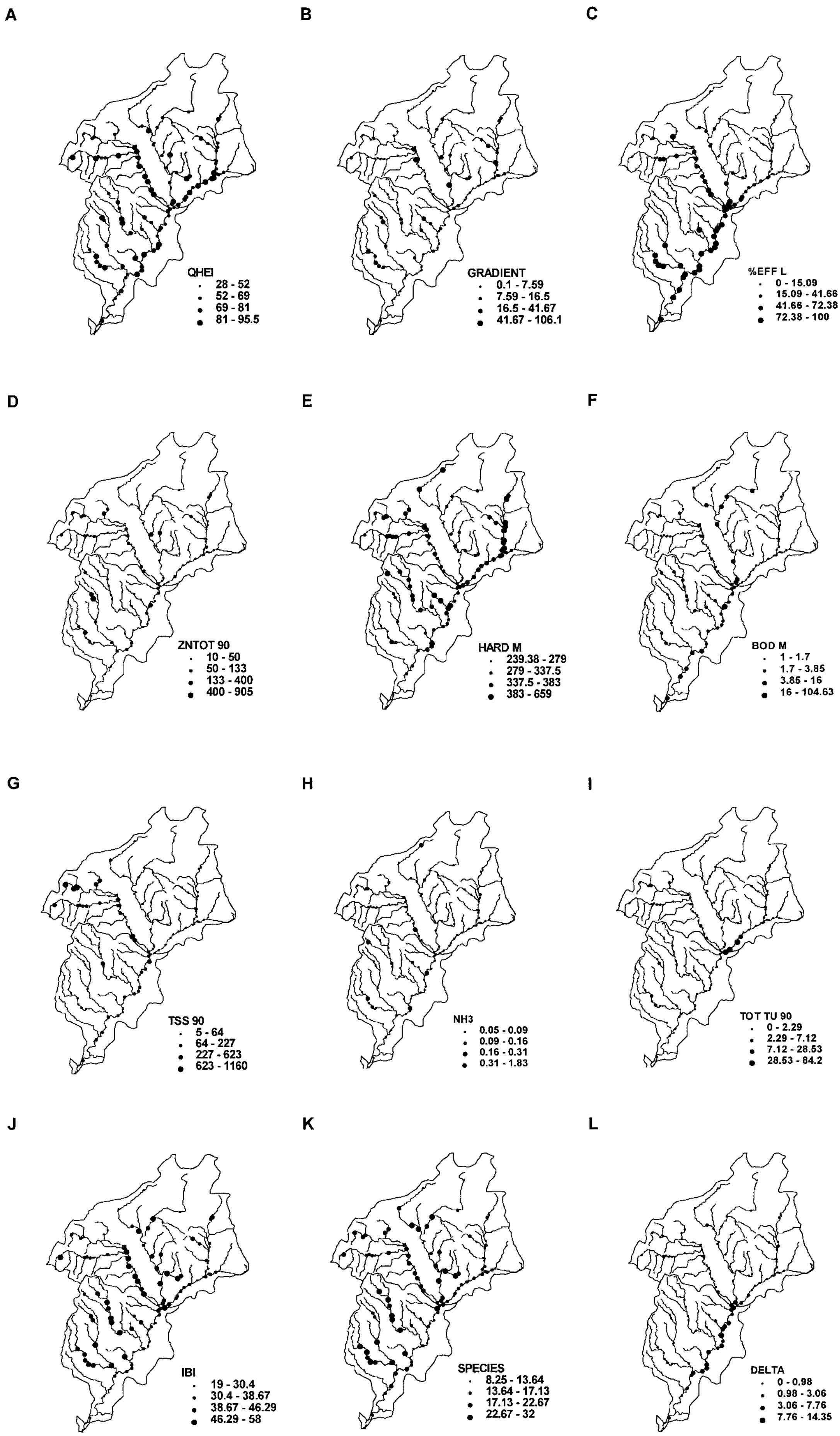


Fig. 2. Designation of subbasins within the Great Miami River, Ohio, USA.

the three best two-variable models, and so on, continuing up to five-variable models. Before the all possible regressions calculations, certain potential regressors were eliminated from consideration because they were too frequently unmeasured, or because 95% or more of their values were equal to the minimum measured value (at detection limit). The complexity of the model was determined by examining the progression of R^2 values as a function of model size (increases in model size were permitted only where R^2 was increased by at least 5%). Up to three models with similar R^2 values were subsequently fit in a forward variable selection process to identify the relative importance of predictors, partial R^2 values, and to allow inspection of the regression coefficients for biological relevance.

Top-down analyses

Study area: state of Ohio. The top-down approach started with a cluster analysis of sites throughout the state of Ohio. The cluster analysis partitioned the sites into groups based on the Euclidean distance between measurements. The second step in the analysis was a regression or analysis of variance to evaluate how factors not used in the clustering process varied across clusters. Three approaches were used in the clustering of sites. First, sites were clustered using habitat information. Variables used in the analysis were QHEI, Drainage Area, Gradient, median hardness, and median pH. Although hardness and pH are truly chemical measures, they were used in this case as parameters that reflect characteristics of the local geology. The data on these variables were first ranked and the ranks from each site were then used to cluster all sites. The number of clusters was varied and then compared to evaluate the number of clusters that discriminates the data best. Stepwise regression analysis was then used to evaluate relationships between the IBI and a number of water chemistry variables. The analysis used the raw variables and log transformed variables. Because of the presence of missing values, a data imputation method was used [12]. The imputation pro-


A
QHEI
28 - 52
52 - 69
69 - 81
81 - 95.5
B
GRADIENT
0.1 - 7.59
7.59 - 16.5
16.5 - 41.67
41.67 - 106.1
C
%EFF L
0 - 15.09
15.09 - 41.66
41.66 - 72.38
72.38 - 100
D
ZNTOT 90
10 - 50
50 - 133
133 - 400
400 - 905
E
HARD M
239.38 - 279
279 - 337.5
337.5 - 383
383 - 659
F
BOD M
1 - 1.7
1.7 - 3.85
3.85 - 16
16 - 104.63
G
TSS 90
5 - 64
64 - 227
227 - 623
623 - 1160
H
NH3
0.05 - 0.09
0.09 - 0.16
0.16 - 0.31
0.31 - 1.83
I
TOT TU 90
0 - 2.29
2.29 - 7.12
7.12 - 28.53
28.53 - 84.2
J
IBI
19 - 30.4
30.4 - 38.67
38.67 - 46.29
46.29 - 58
K
SPECIES
8.25 - 13.64
13.64 - 17.13
17.13 - 22.67
22.67 - 32
L
DELTA
0 - 0.98
0.98 - 3.06
3.06 - 7.76
7.76 - 14.35

cedure ensured that sites with partial data were not eliminated from the analysis.

A second set of cluster analyses was run on the biological metrics that comprise the IBI. Metrics were first clustered to produce groups of sites with similar fish community structure. These clusters were then compared on a number of habitat and water chemistry characteristics. The analysis was based on the analysis of variance of physicochemical variables.

Another approach to grouping sites was by combining basins. In this approach, the basins (Table 2) were first evaluated using correlations between IBI and the habitat variables. Ideally, the basins would be combined based on similar regression relationships. However, missing observations made this approach infeasible. Also, properties of the data such as skewness and outliers precluded the use of ordinary regression analysis. To measure the relationship between habitat and IBI, Kendall's nonparametric correlation coefficients were computed [13]. Kendall's coefficient provides a good measure of monotonic relationships (including linear). The data set was thus reduced from one involving IBI and habitat variables at each site to an array of correlations, with one set of correlations for each basin. The correlation array was also useful for adjusting for missing values. Some basins (3, 6, 7, 8, 10, 12, 21, and 23) did not have sufficient data (number of sites < 10) and were omitted because of insufficient sample size. The basin correlations were then clustered using the average linkage method on SAS (SAS Institute, Cary, NC, USA). This procedure merges basins into groups based on the average Euclidean distance between groups. Groups are formed that have similar correlations between habitat and IBI. Cluster identification is then used as a grouping or class variable to define groups for regression analysis. Stepwise analysis was used to find a set of regression predictors.

RESULTS AND DISCUSSION

Bottom-up approach

Relationships of habitat and ambient water chemistry to selected biological indices and metrics for the entire GMR and identified subbasins were determined via maps and multiple linear regressions. The QHEI and its metrics were used as measures of habitat suitability for a robust and diverse fish community (Fig. 3). The QHEI ranged widely throughout the GMR basin. In general, the poorest habitat index scores were found in the headwater reaches of the various major tributaries. In most situations, modifications due to agricultural practices (channelization, reduced stream cover, reduced riparian area, poorer substrates) were identified as the primary causative factors for reducing the QHEI [6–8]. Poor QHEI scores also were observed in significant stretches of the Lower GMR, downstream of Dayton. With the exception of the Mad River, most streams' habitat scores increased with increased distance from the headwater (i.e., increased drainage area). Between the city of Springfield and the headwater reaches of the Mad River, the QHEI scores were low because of historical channelization and reduced cover [14]. In addition, this stretch of river and the Lower GMR (downstream of Dayton) had little vertical gradient.

Much of the GMR basin is effluent dominated during mean and low flow (Fig. 3). The percent cumulative effluent increased from the headwaters of the Stillwater and Mad rivers to the city of Dayton. Less than 15% effluent at low flow was calculated along the majority of the Upper GMR, Twin Creek, and other tributaries. The influence of effluent is quite large downstream of Dayton, because up to 100% of the flow can be calculated as effluent during low-flow periods. Consequently, water chemistry parameters often associated with effluent discharge, such as biochemiacl oxygen demand (BOD), total suspended solids (TSS), and NH3 were elevated primarily in areas receiving a high percent of cumulative effluent. The number of toxic units (from metals and ammonia) was only elevated in the lowest segments of the Mad River. An inverse relationship apparently existed with Hardness and %Effluent, because Hardness was least in river segments receiving high percentages of effluent. No obvious spatial patterns were observed for Zinc in the watershed.

The greatest IBI scores were found in the Upper GMR, Stillwater River, Twin Creek, and other tributaries. The lowest IBI scores were observed throughout most of the Mad River, headwater regions of the Stillwater and Upper GMR and the entire Lower GMR. About 45% of the variation in IBI scores within the entire GMR could be accounted for in a three-variable model that incorporates QHEI, followed by percent cumulative effluent at mean flow (%Effl M), median Hardness, or Gradient (Table 3). Higher IBI scores were associated with better habitat, whereas effluent had a negative influence. A near doubling of the regression fits of IBI with water chemistry and habitat variables was observed by analyzing the data on a subwatershed basis. About 80% of the variance in IBI within the Lower GMR was accounted for via a five-variable model. In this case, habitat factors such as Pool and Channel scores were the dominant factors, addressing up to 60% of the variance, followed by chemical parameters such as TSS, Hardness, Zinc, Ammonia, and Selenium.

Habitat factors such as Pool, Riffle, and Gradient seemed to be the primary (and positive) ecological drivers for the IBI within the Stillwater, Twin Creek, and Upper GMR, addressing between 45 and 69% of the variation. Ammonia had a negative influence within the Upper GMR and Twin Creek subbasins. The TSS, TOC, and total number of toxic units also had negative relationships with IBI within the Twin Creek subbasin.

Interpretation of the multiple linear regression results from the Mad River was dubious, because Lead and Zinc concentrations and Riffle scores performed opposite to expectations (i.e., if Lead and Zinc were toxicants then a negative relationship should have resulted, as also a positive relationship with good riffle habitat should have yielded a more diverse fish community, hence an increased IBI).

A similar spatial pattern of the number of fish taxa (Species) to that of the IBI thoughout the GMR was observed. That is, the greatest species richness occurred in the Upper GMR, Twin Creek, and other tributaries. The Mad River and the Lower GMR contained the least number of fish taxa. Only about 30% of the variance in Species could be accounted for within the entire GMR, using a three-variable model incorporating Hardness, Pool, and %Effluent at mean and low flows. A positive relationship existed with Pool score (as with QHEI and IBI)

←

Fig. 3. Spatial distribution of key habitat, chemical, and fish community variables within the Great Miami River, Ohio, USA. The size of the dot is proportional to the amount of the variable.

Table 3. Summary of stepwise multiple linear regressions for the Great Miami River (GMR) and subbasins (Lower GMR, Mad River, Twin Creek and other tributaries, Stillwater River, and Upper GMR). Values represent cumulative coefficients of determination per step[a]

Dependent variable	Watershed	No.	Step 1	Step 2	Step 3	Step 4	Step 5
IBI	GMR	1	QHEI 0.17	−%Effl M 0.40	−Hard M 0.46		
IBI	GMR	2	QHEI 0.17	−%Effi M 0.40	Gradient 0.44		
IBI	Lower	1	Pbtot 90 0.21	Pool 0.33	−Channel 0.68	Zntot M 0.77	−TSS 90 0.82
IBI	Lower	2	Pool 0.13	−Channel 0.54	Zntot M 0.69	Hard 90 0.70	Hard M 0.82
IBI	Lower	3	Pool 0.17	−Channel 0.60	Zntot M 0.62	−NH3 M 0.67	Setot 90 0.81
IBI	Mad	1	Pbtot M 0.29	−Riffle 0.54	Zntot M 0.69	−Hard 90 0.80	
IBI	Stillwater	1	Pool 0.55	DO 90 0.71	−Hard M 0.78		
IBI	Twin and other creeks	1	Pool 0.45	−NH3 90 0.69	BOD 90 0.76	−TSS M 0.81	
IBI	Twin and other creeks	2	Pool 0.45	−TOC 90 0.56	−TotTUs M 0.66	BOD 90 0.80	
IBI	Upper	1	Riffle 0.58	Gradient 0.69	−Nitot M 0.76	−NH3 M 0.82	
Species	GMR	1	−Hard M 0.12	Pool 0.20	−%Effl M 0.30		
Species	GMR	2	Pool 0.08	−%Effl M 0.20	%EfflL 0.29		
Species	GMR	3	Hard M 0.12	QHEI 0.18	%Effl M 0.27		
Species	Lower	1	−Channel 0.08	Pool 0.32	−NH3 90 0.41	Zntot 90 0.55	Setot M 0.67
Species	Lower	2	−Riparian 0.08	−NH3 90 0.13	Zntot 90 0.26	Hard 90 0.43	Gradient 0.65
Species	Lower	3	−Riparian 0.08	−TSS M 0.11	−Hard M 0.16	Setot M 0.53	−Drainage Area 0.64
Species	Mad	1	Pbtot M 0.34	−NH3 M 0.48	Substrate 0.60	−Hard M 0.68	
Species	Stillwater	1	Pool 0.25	Cdtot 90 0.45	DO M 0.59	TOC 90 0.74	
Species	Stillwater	2	Pool 0.25	TOC 90 0.44	DO M 0.63	Zntot M 0.72	
Species	Stillwater	3	Pool 0.25	Cdtot 90 0.45	DO M 0.59	BOD 90 0.70	
Species	Twin and other creeks	1	Pool 0.53	−Zntot M 0.69	BOD 90 0.77		
Species	Twin and other creeks	2	Pool 0.53	−Zntot M 0.69	−QHEI 0.75		
Species	Upper	1	Channel 0.26	Cover 0.37	−PbtotM 0.48	−Nitot M 0.55	−Drainage Area 0.66
DELTs	GMR	1	%Effl M 0.44				
DELTs	Lower	1	−TSS M 0.10	NH3 M 0.20	−Hard 90 0.39	−QHEI 0.47	Channel 0.78
DELTs	Mad	1	BOD 90 0.41	−Setot 90 0.57	Substrate 0.75	Drainage Area 0.78	−QHEI 0.88
DELTs	Stillwater	1	Pbtot M 0.27	Substrate 0.43	−QHEI 0.45	Cover 0.60	
DELTs	Twin and other creeks	1	NH3 M 0.20	−TOC M 0.40	−Zntot 90 0.54	−Pool 0.69	Tot TUs 90 0.79
DELTs	Twin and other creeks	2	NH3 M 0.20	−TOC M 0.40	−Riparian 0.64	−Zntot 90 0.73	−TSS 90 0.79
DELTs	Upper	1	TSS M 0.31	−DO M 0.53	Riffle 0.62	−TSS 90 0.68	−Riparian 0.73

[a] Parameter names are spelled out in Table 1.

and a negative relationship was found with %Effluent (also as with IBI). In the Lower GMR, 64% of the variance in Species (richness) was accounted for via a five-variable model that included Riparian, TSS, Hardness, Selenium, and Drainage Area. All but median Selenium concentrations were negatively related to Species. A four-variable model best fit the Species within the Mad River, where an overall coefficient of determination of 0.68 was observed, including the parameters Lead,

Table 4. Means of habitat variables used to derive habitat-based clusters

Parameter	Cluster 1	2	3	4
QHEI[a]	66.51	74.96	51.62	46.75
Drainage Area	29.23	692.57	759.90	3,193.55
Gradient	17.07	6.74	3.10	13.94
Hardness	286.61	288.65	293.66	145.50
pH	7.85	7.99	7.86	7.45
Sample size	149	237	208	1

[a] QHEI = qualitative habitat evaluation index.

Table 5. Summary of stepwise regression results for models relating biological and chemical variables for three habitat-based clusters. Regression coefficients are in parentheses[a]

Response	Cluster 1	Cluster 2	Cluster 3
IBI	%EffM (−0.5)	%EfflM (−0.48)	DO (0.88)
	NH3 (−53.3)	DO (0.92)	Total P (−5.13)
	Cd total (−8.4)		%EfflM (−0.24)
			Total Zn (0.005)
	$R^2 = 0.2795$	$R^2 = 0.2291$	$R^2 = 0.2515$

[a] Parameter names are spelled out in Table 1.

Ammonia, Substrate, and Hardness. For the other three subbasins, Stillwater, Twin Creek, and Upper GMR, the first significant step was a habitat variable (e.g., Pool, Channel, or Cover). Further positive relationships of Species with water chemistry parameters (Cd, TOC, DO, Zn) were observed for the Stillwater River. The negative correlation with Zinc in the Twin Creek subbasin should be considered with caution because it is driven by elevated concentrations at 2 of 19 segments evaluated.

The spatial distribution of the percent of fish observed with DELTs within the entire GMR was not as observed for the other two fish parameters. In this case, DELTs were only elevated within the city of Dayton (lowest segments of the Mad and Upper GMR) and in the Lower GMR. Water chemistry parameters accounted for the greatest variance in DELTs within the entire GMR as well as subbasins, unlike that observed for IBI and Species. Forty-four percent of the DELT variation was accounted by %Effluent at mean flow for the entire GMR. With exception to the Lower GMR, the variation of DELTs within the other four subbasins was small, hence interpretation of the multiple regressions should be considered with caution. The DELTs within the lower GMR were found to be associated with TSS, Ammonia, Hardness, QHEI, and Channel.

Although the purpose of this exercise was to illustrate the utility of the bottom-up approach in understanding the potential effects of multiple stressors on receiving water biology, it was not intended to be a comprehensive assessment of all the potential effects caused by diverse stressors within the GMR, such as habitat alterations, toxic effects from combined sewer overflows, effluent discharge, sediments, and so on. The purpose of exploring this approach was to determine the relationships of measured habitat and chemical stressors on the biota and explore the effects of scale (basin vs subbasin) on the analysis. The principle statistical tool for assaying the strength of various stressor–response relationships was regression analysis. The best fit regressions were determined for the subbasin level. In most cases, between 60 and 80% of the variance of the response variables was addressed with three to five variables at the subbasin level, whereas less than 30% was commonly fit when analyzing the entire GMR. Even so, caution should be exercised so as to not overinterpret the regression results. As discussed before, the variance of %Effluent in the Lower GMR was low, and, therefore, did not appear as a significant contributor in the depression of IBI and Species. However, examination of the map of each of these dependent variables and DELTs clearly indicated that a relationship exists with effluent.

The utility of the bottom-up approach is dependent on analysis of more than one response variable (e.g., IBI), but a set of variables that can be used together to diagnose effects from multiple stressors. For instance, toxic influences would be expected to depress the number of fish taxa observed as well as IBI, whereas the percent of DELTs may increase. However, where DELTs do not respond in an inverse manner to that of species richness and IBI, toxicity as a prime stressor may be doubted. Depressed species richness as well as elevated DELTs occurred downstream of Dayton (Lower GMR), where percent cumulative effluent was the greatest. Although this may seem to point to municipal effluents as the primary causative agents of adverse effects in the GMR, this conclusion may not be warranted because significant percentages of effluent were also determined for major portions of the Stillwater River and Twin Creek subbasins, where IBI and Species values were among the highest recorded in the state of Ohio [8], whereas DELTs were minimized. Hence, the adverse effects observed downstream from Dayton may be due to the discharge of contaminants unknown within the contexts of this study (i.e., more than just municipal effluent), or possibly from the joint action of effluent discharge and habitat modification (channelization, impoundments) [8]. If the latter, regression modeling provides an estimate of this joint action.

We believe that percent cumulative effluent may be useful for determining stressor source–biological response relationships; however, its best use may be as a screening tool, or surrogate parameter, that encompasses unknown toxics (as single compounds or as a mixture), nutrient effects or flow modifications. Determining the agents in effluents potentially responsible for adverse effects within the receiving water environment may require additional data inputs, such as sediment chemistry, whole-effluent toxicity test results, and body burdens of suspected chemicals within the receiving water biota.

Top-down approach

Habitat clusters. The cluster analysis of the habitat data resulted in three moderate-size clusters and a single site. The clusters are summarized in Table 4. Cluster 1 represented sites characteristic of headwaters, for example, small drainage areas and high gradients with moderately good habitat. Clusters 2 and 3 corresponded to sites with larger drainage area but differed in overall habitat and gradient, with cluster 2 having the greater mean QHEI score and greater gradients. Hardness and pH did not contribute toward the separation of clusters 1 through 3; however, they were important in the identification of cluster 4, where both hardness and pH levels were lower compared to the other three clusters. Importantly, cluster 4 reflected a single site with a very large drainage area. Stepwise regression analyses are summarized in Table 5. Overall, the regressions were weak, with approximately 25% of the variance of IBI being addressed by water chemistry variables with-

Table 6. Summary of stepwise regressions using imputed covariance matrices for each of the habitat-based clusters[a]

Dependent variable	Cluster	Step 1	Step 2	Step 3	Step 4
IBI	1	−BOD 0.17	−%Effl M 0.23	−Zinc 0.26	
IBI	2	−Effl M 0.15	DO M 0.19	−BOD 0.23	
IBI	3	−Zinc 0.15	DO 90 0.30	−%Effl L 0.37	−TSS 90 0.41

[a] Parameter names are spelled out in Table 1.

Table 8. Summary table on separation of noncluster variables from biologically based clusters. The greater the *F* statistic, the greater the separation due to clustering[a]

F statistics			
86.22–40.91	27.45–12.45	8.90–3.1	2.6–1.17
QHEI	%Effl M	pH M	Hard M
Drainage Area	DO M	Zntot 90	TotTU 90
Gradient	TSS M	%Effl L	Cdtot 90
	TotP M	Zntot M	TotTU M
	NH3 M	TSS 90	
		Cutot M	

[a] Parameter names are spelled out in Table 1.

in each of the three main clusters. Across each cluster, the percent cumulative effluent at mean flow was consistently identified as a negative influence on IBI. In headwater streams, cluster 1, ammonia and total cadmium were also identified as negative influences on IBI. Dissolved oxygen was a positive influence within both clusters 2 and 3. When imputed covariance matrices were used as inputs for the stepwise regressions, BOD became more important in sites with small drainage areas (cluster 1), followed by percent effluent at mean flow (Table 6). Effects of imputation were not observed for cluster 2, whereas zinc, effluent at low flow, and TSS along with DO 90 addressed 41&percnt: of IBI variation in cluster 3.

Biological clusters. Clusters of fish metric data were used to ascertain the importance of physical and chemical characteristics on the separation of the clusters. Three clusters were developed from nine IBI metrics (Table 7). The results suggest that cluster 1 is associated with poor fish community integrity, exemplified by a majority (60%) of the taxa sampled being identified as tolerant species. Further indications of poor biotic integrity include elevated DELTs, high percentage of omnivores (Omnivor), low round-bodied sucker (Rdsuckpc) composition, and very low numbers of intolerant fish taxa (Intols). More robust fish communities correspond to cluster 2, where greater numbers of intolerant taxa and lower percentages of tolerant taxa, omnivores, and DELTs were present. Cluster 3 had a higher mean of Intols and Darters, but fewer numbers of sucker species and round-bodied suckers than in cluster 2. When evaluated on the variables not used in the cluster analysis (Table 8), the greatest differences were associated with habitat variables (QHEI, drainage area, and gradient) followed by weaker differences due to some of the chemical variables (percent effluent at mean flow, median DO, TSS, total phosphorus). Hence, to this point, the two different clustering approaches (habitat- and fish metric-based) have yielded similar interpretations—that although water chemistry and percent effluent may account for a significant amount of the variation in IBI throughout the state, the dominating factors seem to be more oriented to habitat, including drainage area and gradient, with water chemistry playing a secondary role.

Basin clusters. Clusters based on the correlations of IBI to QHEI, Drainage area, median Hardness M, and median pH were determined for basins that had a minimum of 10 observations per basin. Three clusters were determined at a Euclidean distance of 0.9 (Fig. 4): cluster 1 (1, 2, 4, 5, 9, 11, 17), cluster 2 (14, 15, 19), and cluster 3 (13, 16, 18, 20). In the first cluster, QHEI and pH were correlated most strongly with IBI, yielding correlation coefficients of 0.38 and 0.31, respectively. In cluster 2, drainage area ($r = -0.37$) and median Hardness ($r = -0.37$) were most strongly correlated, whereas in cluster 3 all variables seemed to be correlated with IBI: QHEI ($r = 0.28$), drainage area ($r = -0.35$), hardness ($r = 0.47$), and pH ($r = 0.34$). Figure 5 shows where each of the basins corresponding to the three clusters occurred. Cluster 1 covered the largest geographical area in the state with basins located in the northwestern, central, and southeastern parts of Ohio. Although three basins were identified in cluster 2, they corresponded to two areas that drain the metroplexes of greater Dayton–Cincinnati (GMR) and Cleveland (Chagrin River and

Table 7. Means of fish metrics used to derive biologically based clusters

Metric	Cluster 1 (mean)	Cluster 2 (mean)	Cluster 3 (mean)
Sunfish (no. sunfish species)	2.07	3.54	2.73
Darters (no. darter species)	0.95	2.12	3.50
Suckers (no. sucker species)	1.04	4.06	2.33
Intols (no. intolerant species)	0.07	1.75	1.12
Tolperc (percent tolerant species)	60.80	23.31	38.86
Topcarn (percent top carnivores)	3.03	8.87	3.99
Omnivor (percent omnivores)	31.20	24.79	19.68
Rdsuckpc (percent round-bodied suckers)	0.43	16.72	4.25
DELTs (percent deformities, fin erosions, lesions, and tumors)	3.04	2.15	0.36
N	281.00	538.00	356.00

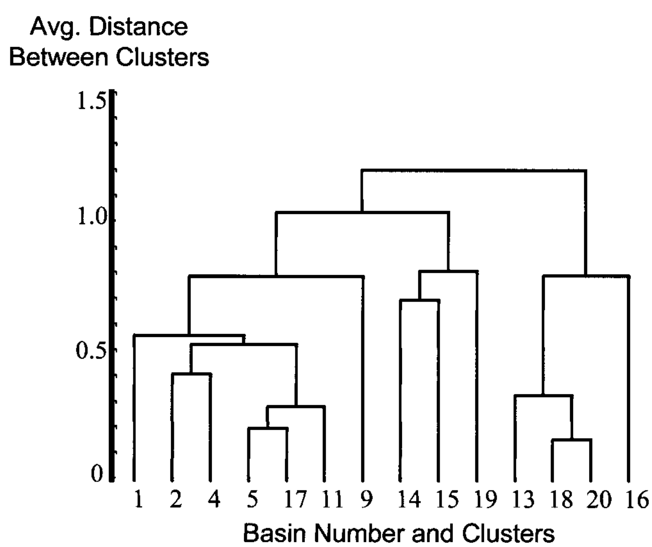


Fig. 4. Average Euclidean distance between clusters for basins for which sufficient data were available (i.e., greater than 10 sites).

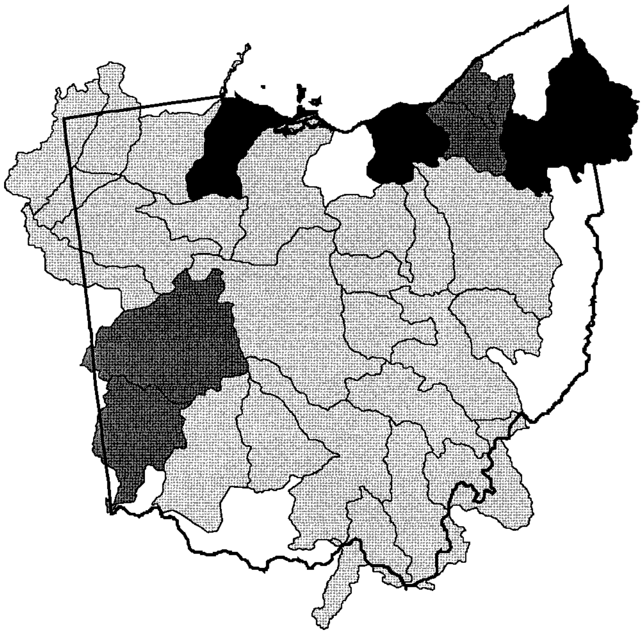


Fig. 5. Location of basin clusters based on correlation matrix of the index of biotic integrity versus habitat parameters.

Cuyahoga River). The third cluster corresponded to four rivers within the Lake Erie drainage in the northern portion of the state. Stepwise regressions of habitat and water chemistry variables to IBI within each basin cluster were determined. In cluster 1, a three-variable model fit 39% of the variation in IBI, with BOD M and QHEI contributing 35%. Percent effluent at mean flow and QHEI addressed 34% of the variance in cluster 2, whereas a combination of QHEI, BOD M, and DO 90 fit 63% of the variation in IBI in cluster 3. Across each cluster, QHEI and DO 90 were positively correlated with IBI, whereas %Effl M and BOD M were negatively related to IBI.

A comparison of results from the various top-down approaches with the bottom-up approach indicated that as the geographical dispersion of sampling sites increased (scale increased), the degree of variation in IBI fit by water chemistry and/or habitat data decreased. In general, between 25 and 40% of the variation of IBI could be fit to any set of habitat and/or water chemistry variables using habitat-based, biologically based, and basin-based clustering methods. This was about the same goodness of fit as was found for IBI in the GMR (entire basin) via the bottom-up approach.

CONCLUSIONS

Considering all the analyses conducted, several consistencies were found: quality of habitat, expressed as QHEI, Gradient, and Drainage Area, were generally observed as positive influences on IBI; in contrast, high percent cumulative effluent was generally a negative influence on the overall IBI score, with conventional pollutants such as ammonia and BOD also negatively correlated with IBI; and metals and total toxic units, although often found in significant regressions, were inconsistently related to IBI or other fish metrics and, therefore, do not seem to be significant drivers of IBI over a large geographical area, although they may be important within the subbasin level.

Further analyses of the Ohio data sets are ongoing. At present, we plan on continuing the bottom-up approach for each basin in the state using the described river segmentation approach with the hopes of ascertaining fish and invertebrate community–response signatures that may be clustered using significant regression coefficients. These clusters can then be compared to those briefly described in this paper. We have also developed a spatially driven data imputation methodology currently under evaluation that will reduce many of the problems associated with missing values, as described in this manuscript.

Acknowledgement—We wish to thank Bob Miltner, Dennis Mischne, Ed Rankin, Dale White, and Chris Yoder of the Ohio Environmental Protection Agency for supplying water chemistry, fish community, and habitat data as well as lending their expertise in identifying areas of future method exploration and development.

REFERENCES

1. U.S. Environmental Protection Agency, Region IX. 1992. STORET documentation for menu-driven user interface. EPA 68-C9-0013. San Francisco, CA.
2. Dyer SD, White-Hull CE, Wang X, Johnson TD, Carr GJ. 1998. Determining the influence of habitat and chemical factors on instream biotic integrity for a southern Ohio watershed. *J Aquat Ecosys Stress Recov* 6:91–110.
3. Rankin ET. 1989. The qualitative habitat evaluation index (QHEI): Rationale, methods and application. Division of Water Quality Planning and Assessment, Columbus, OH, USA.
4. Ohio Environmental Protection Agency. 1988. *Biological Criteria for The Protection of Aquatic Life,* Vol 2—Users Manual for Biological Field Assessment of Ohio Surface Waters. Columbus, OH, USA.
5. U.S. Environmental Protection Agency. 1992. Environmental and program information systems compendium. EPA 800-B92-001. Washington, DC.
6. U.S. Environmental Protection Agency. 1989. 1988 Needs Survey: Report to Congress—Assessment of needed publicly owned wastewater treatment facilities in the United States, February, 1989 EPA 430/9-89-001. Washington, DC.
7. U.S. Environmental Protection Agency. 1992. Permit compliance system user documentation master index. PCS-MA93-1.01. Washington, DC.
8. Ohio Environmental Protection Agency. 1997. Biological and water quality study of the Middle and Lower Great Miami River and selected tributaries, 1995, Vol I. Ohio EPA Technical Report MAS/1996-12-8. Columbus, OH, USA.
9. Ohio Environmental Protection Agency. 1996. Biological and water quality study of the Upper Great Miami River and selected tributaries. Ohio EPA Technical Report MAS/1995-12-13. Columbus, OH, USA.
10. Ohio Environmental Protection Agency. 1997. Biological and water quality study of Twin Creek, 1995. Ohio EPA Technical Report MAS/1996-12-9. Columbus, OH, USA.
11. Hocking RR. 1976. The analysis and selection of variables in linear regression. *Biometrics* 32:1–50.
12. Schafer JL. 1997. *Analysis of Incomplete Multivariate Data.* Chapman and Hall, London, UK.
13. Helsel DR, Hirsch RM. 1992. *Statistical Methods in Water Resources.* Elsevier, New York, NY, USA.
14. Ohio Environmental Protection Agency. 1994. Biological, sediment, and water quality study of the Lower Mad River and Hebble Creek, Dayton, Ohio. Technical Report EAS/1994-6-9. Columbus, OH, USA.